AFRL-SA-WP-SR-2016-0020

Automated Vision Test Development and Validation

Steve Wright, Darrell Rousse, Alex van Atta, James Gaska, Marc Winterbottom, Steven Hadley, Lt Col Dan Lamothe

November 2016

Air Force Research Laboratory
711th Human Performance Wing
U.S. Air Force School of Aerospace Medicine
Aerospace Medicine Department
2510 Fifth St., Bldg. 840
Wright-Patterson AFB, OH 45433-7913

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-SA-WP-SR-2016-0020 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNATURE//	//SIGNATURE//
_____	_____
DR. DANIEL L. VAN SYOC	COL PATRICK R. STORMS
Deputy Chief, Aerosp Med Consultation Div	Chair, Aerospace Medicine Department

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
7 Nov 2016	Special Report	April 2014 – April 2015

4. TITLE AND SUBTITLE

Automated Vision Test Development and Validation

5a. CONTRACT NUMBER
FA8650-10-D-6056

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steve Wright, Darrell Rousse, Alex van Atta, James Gaska, Marc Winterbottom, Steven Hadley, Lt Col Dan Lamothe

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
USAF School of Aerospace Medicine
Aerospace Medicine Department/FECO
2510 Fifth St., Bldg. 840
Wright-Patterson AFB, OH 45433-7913

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-SA-WP-SR-2016-0020

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES
Cleared, 88PA, Case # 2016-6017, 23 Nov 2016.

14. ABSTRACT

The Optec Vision Test, originally produced in 1951 as the Armed Forces Vision Tester, is the sole device used to qualify individuals for Air Force flight duties. Although the external appearance of the device has changed since its first inception, the design of the slides used to present the visual stimuli is exactly the same as those originally produced. The goals of this effort were to evaluate proof of concept that the vision screening tests currently administered using the Optec Vision Test could be transitioned to a computer-based, automated system; to produce software for desktop displays; and to evaluate features such as user interfaces, threshold algorithms, validity of results, and screening techniques that could minimize testing time. This was a prospective study consisting of 27 individuals aged 18-40, a range that represents the Air Force flying population. There was no stated requirement for gender and there were no exclusion criteria related to visual status as subjects with both normal and non-normal visual skills were desired. Automated, computer-based vision tests were developed to assess high and low (5% Michelson) contrast visual acuity, letter contrast sensitivity at 20/25 and 20/50 acuity levels, color contrast sensitivity, and stereoacuity. The current effort demonstrates that automated vision tests produce reliable results, with coefficients of determination for repeated testing above 0.80 for many of the tasks. However, to achieve this high level of reproducibility and reduce the standard error of the threshold estimate to near asymptotic levels, 30 or more trials are often required. Successful implementation of automated (or any) vision testing in an aerospace medicine clinic requires methods of determining visual status quickly, but with high accuracy. We found that 100% sensitivity could be achieved using a fast screening method, however, at the cost of performing full threshold testing on over 30% of normal subjects, which is quite time consuming. This effort was accomplished using desktop monitors; however, future efforts will pursue transitioning these tests to a system designed to standardize test distance and illumination conditions and eliminate the potential for head movements, and with a form factor suitable for more routine clinical use.

15. SUBJECT TERMS
Vision screening, visual acuity, contrast sensitivity, stereoacuity, computer-based, automated testing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Steven Wright
U	U	U	SAR	19	19b. TELEPHONE NUMBER *(include area code)*

This page intentionally left blank.

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

Page

LIST OF TABLES

Page

1.0 SUMMARY

The Optec Vision Test, originally produced in 1951 as the Armed Forces Vision Tester, is the sole device used to qualify individuals for Air Force flight duties. Although the external appearance of the device has changed since its first inception, the design of the slides used to present the visual stimuli is exactly the same as those originally produced. The goals of this effort were to evaluate proof of concept that the vision screening tests currently administered using the Optec Vision Test could be transitioned to a computer-based, automated system; to produce software for desktop displays; and to evaluate features such as user interfaces, threshold algorithms, validity of results, and screening techniques that could minimize testing time. This was a prospective study consisting of 27 individuals aged 18-40, a range that represents the Air Force flying population. There was no stated requirement for gender and there were no exclusion criteria related to visual status as subjects with both normal and non-normal visual skills were desired. Automated, computer-based vision tests were developed to assess high and low (5% Michelson) contrast visual acuity, letter contrast sensitivity at 20/25 and 20/50 acuity levels, color contrast sensitivity, and stereoacuity. The current effort demonstrates that automated vision tests produce reliable results, with coefficients of determination for repeated testing above 0.80 for many of the tasks. However, to achieve this high level of reproducibility and reduce the standard error of the threshold estimate to near asymptotic levels, 30 or more trials are often required. Successful implementation of automated (or any) vision testing in an aerospace medicine clinic requires methods of determining visual status quickly, but with high accuracy. We found that 100% sensitivity could be achieved using a fast screening method, however, at the cost of performing full threshold testing on over 30% of normal subjects, which is quite time consuming. This effort was accomplished using desktop monitors; however, future efforts will pursue transitioning these tests to a system designed to standardize test distance and illumination conditions and eliminate the potential for head movements, and with a form factor suitable for more routine clinical use.

2.0 BACKGROUND

"Present military visual standards have existed with little real change since WWII. The design of instruments used to measure visual acuity (VA), color vision, and muscle balance in military clinical settings remains unchanged since the original purchases over 40 years ago." Since Moffitt and Genco made that statement over 25 years ago, military vision screening tests have remained essentially unchanged and continue to rely on World War II era technology [1].

Many current military vision standards were established by the Armed Forces National Research Council Vision Committee from 1944 to 1954 [2]. The committee consisted of physicians and scientists representing the three military branches as well as academia. They met several dozen times at various locations across the United States and proposed standards for a wide range of visual attributes including color vision, VA, heterophoria, and depth perception. In addition to establishing standards, they further developed the specific tests used to measure these attributes as well as the specific device that would be used to administer the tests. This device, then called the Armed Forces Vision Tester, was originally produced by Bausch and Lomb (Bridgewater, NJ) in 1951 (Figure 1). It is now marketed by Stereo Optical (Chicago, IL) as the Optec 2300 or Optec Vision Test (OVT) (Figure 2) and is the sole device used to qualify individuals for U.S. Air Force (USAF) flight duties. Although the external appearance of the

device has changed since its first inception, the design of the slides used to present the visual stimuli is exactly the same as those originally produced.

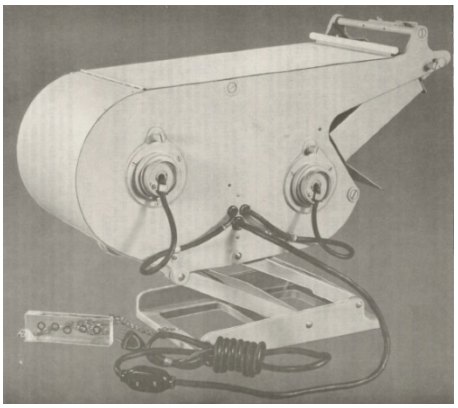

Figure 1. Original design of Armed Forces Vision Test, circa 1951.

Figure 2. Current design of Optec 2300, 2015.

The OVT provides a highly effective medium for vision screening. The all-inclusive "box" design ensures standardization of test distance and illumination conditions. However, the OVT utilizes visual stimuli imprinted on transparencies sandwiched between two glass slides, precluding the ability to randomize or modify test presentation. Thus, the Snellen letters used to qualify a person for flight duties are the same year after year, introducing the risk of memorization. A more extreme example of potential test compromise is the fact that an individual could purchase the entire unit, including slides and answer key, online. Although the manufacturer limits sales to qualified military clinics, the device can be found on a number of sites that sell used medical equipment.

The goal of this effort was to evaluate proof of concept that the vision screening tests currently administered using the OVT could be transitioned to a computer-based, automated system. It was not an attempt to produce an automated vision test, as that would entail developing a manufacturing process. Rather, the goal was to produce software for desktop displays and evaluate features such as user interfaces, threshold algorithms, validity of results, and screening techniques that could minimize testing time. Development of an automated vision test with an industry partner was intended to be a follow-on project.

3.0 METHODS

3.1 Participants

This prospective study was approved by the Air Force Research Laboratory's Wright Site Institutional Review Board (IRB # FWR20140079H). All subjects provided informed consent prior to participation and were free to withdraw at any point during the study. Study participants consisted of 27 individuals aged 18-40, a range that represents the USAF flying population. There was no stated requirement for gender, and there were no exclusion criteria related to visual status, as subjects with both normal and non-normal visual skills were desired.

3.2 Visual Tasks

Automated, computer-based vision tests were developed to assess high and low (5% Michelson) contrast VA, letter contrast sensitivity (CS) at 20/25 and 20/50 acuity levels, color CS, and stereoacuity. These particular attributes were chosen as they represent vision tests currently used as part of the initial and annual vision screenings for USAF aviators (high contrast acuity, color CS, and stereoacuity), are used as a part of the waiver criteria for aircrew after refractive surgery (5% contrast VA), or represent tests administered at the Aeromedical Consultation Service (letter CS at 20/25 and 20/50 acuity levels). All tasks used a four-alternative forced choice (up, down, left, and right buttons) with responses captured on a hand-held keypad. An eight-alternative forced choice was evaluated; however, users found it difficult to select the diagonal responses without shifting their gaze to the keypad. Similarly, voice recognition proved to be unreliable, despite a very limited library of recognizable responses. Given that the results of these tests could have significant impact on an aviator's career, this was deemed unacceptable.

Each task used both a Bayesian adaptive procedure [3-5] to determine true visual threshold as well as a screening mode that would be applicable for routine clinical use. Further details on these will be discussed later. Subjects were tested monocularly (eye selected at random) on all tests with the exception of stereoacuity, and all testing was performed using habitual correction. Each computer-based task was performed twice, using the same eye each time, to assess repeatability characteristics. Test-retest repeatability was not assessed for the chart-based tests that were used for comparison, since memorization could contaminate the results. The randomization possible with computer-based tests is clearly a significant advantage.

The visual stimulus used for acuity, contrast, and color testing was a Landolt C (Figure 3), with the gap oriented at the top, bottom, right, or left position. In all cases, except color, the stimulus was visible for 8 seconds and testing did not continue until a response was offered. Test images were generated using an Intel NUC processor, displayed on a 23-inch liquid crystal display monitor (NEC Multisync, P232W) at 1920x1080 resolution. Proper calibration was confirmed using a spot photometer/colorimeter (X-Rite i1 Display Pro, OEM model). A more detailed description of the color calibration procedure is provided elsewhere [6]. Due to the fact that many of the images were presented at threshold or near threshold levels, an auditory signal indicated when an image was being presented. For similar reasons, peripheral cues (similar to cross hairs) provided an aid to the location of the stimulus.

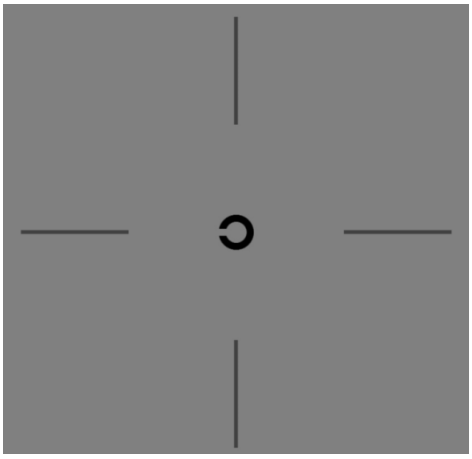

Figure 3. Stimulus used for VA, letter CS, and color CS.

High and low contrast VA and letter CS were correlated with analogous eyecharts currently used by the USAF (Precision Vision, La Salle, IL, SKU 2102, 2186, 2126, and 2128). Testing for both the computer-based tests as well as the eyecharts was accomplished at 4 meters in an otherwise darkened room. When tested on the charts, subjects were instructed to identify as many letters as possible without penalty for errors. LogMAR acuities were determined based on

the formula (42 - # letters correct)*0.02, while log CS was calculated as (0 - # letters correct)*0.05.

Color CS results were correlated with findings from the Rabin cone contrast test, or RCCT (Innova Systems, Hinsdale, IL), which is the standard color vision test used for screening USAF aircrew [7,8]. Our color test, the Operational Based Vision Assessment cone contrast test (OCCT), was similar to the RCCT in that both measure CS while selectively stimulating each of the three retinal cone pigments. However, there were several significant differences between the two tests as follows:

- The RCCT presents stimuli at five fixed contrast levels, while the OCCT presents stimuli at any contrast determined by the adaptive algorithm.
- The RCCT tests the L, M, and S cones in consecutive fashion, while the OCCT test interleaved the colors.
- The RCCT uses a 20/300 letter size for L and M cones and a 20/400 size for the S cone, while the OCCT version used a constant 20/330 stimulus size.
- The RCCT is designed for testing at 36 inches, while the OCCT test was calibrated for 1 meter.
- The RCCT presents the stimulus for 4 seconds with a 400-ms delay before the next image is displayed (regardless of whether a response is offered), while the OCCT presented the image for 3 seconds with a 1,250-ms delay between presentations.

Stereoscopic images were generated with a computer using an Intel Core i7 central processing unit and a NVIDIA GeForce GTX 680 graphics card. Images were displayed on a 27-inch monitor (Asus VG278) with a frame rate of 120 Hz at 1920x1080 resolution (approximately 81 dpi). Isolated visual input to the right and left eye was achieved using liquid crystal display shuttered glasses (NVIDIA 3D Vision 2). The stereo target (Figure 4) was a set of four circles arranged in a diamond pattern. Each circle measured 5 mm (17.2 arcmin) in diameter, while the reference mask measured 30 mm (103.1 arcmin) horizontally and vertically. Testing was accomplished at 1 meter in an otherwise darkened room. At this distance each pixel represented 0.314 mm in size, and if the image was presented based on whole pixel steps, the display would have been limited to testing no better than 65 arcsec. To overcome this limitation and make the test eye limited, anti-aliasing techniques were used [9,10]. This allowed us to accurately measure stereoacuity thresholds of better than 10 arcsec. Results were correlated with a Titmus stereoacuity book (Stereo Optical, Chicago, IL). When administered at 16 inches, the Titmus book will measure down to only 40 arcsec of stereoacuity. To allow comparison to the electronic test, we administered the Titmus test at 1 meter, which allowed for measurement down to 16 arcsec.

Figure 4. Stimulus used for stereoacuity.

4.0 RESULTS

Test-retest characteristics (coefficient of determination, R^2) for each computer-based automated task are reported in Table 1. Each row corresponds to the repeatability if the test was stopped at the given number of trials reported in the first column. Figure 5 provides a graphical representation of the standard error of the threshold estimate based on the number of trials completed for three of the visual tasks. It is evident that to achieve a high level of repeatability (and thus reliability) and to approach the asymptote for error, 30 or more trials are needed for many of the tasks.

Table 1. R^2 for Repeated Trials on Each Task

Trial	VA		CS		Color Vision			Stereoacuity
	High Contrast	Low Contrast	20/50 Letter	20/25 Letter	M Cone	L Cone	S Cone	
5	0.256	0.306	0.337	0.184	0.421	0.199	0.005	0.433
10	0.115	0.757	0.711	0.430	0.654	0.498	0.385	0.489
15	0.439	0.612	0.797	0.497	0.728	0.658	0.465	0.377
20	0.610	0.835	0.765	0.701	0.746	0.792	0.578	0.537
25	0.637	0.858	0.785	0.776	0.824	0.855	0.595	0.567
30	0.622	0.862	0.771	0.809	0.898	0.847	0.601	0.722
35	0.699	0.866	0.818	0.817	0.853	0.852	0.738	NT
40	0.721	0.896	0.826	0.855	0.938	0.883	0.732	NT

NT = not tested.

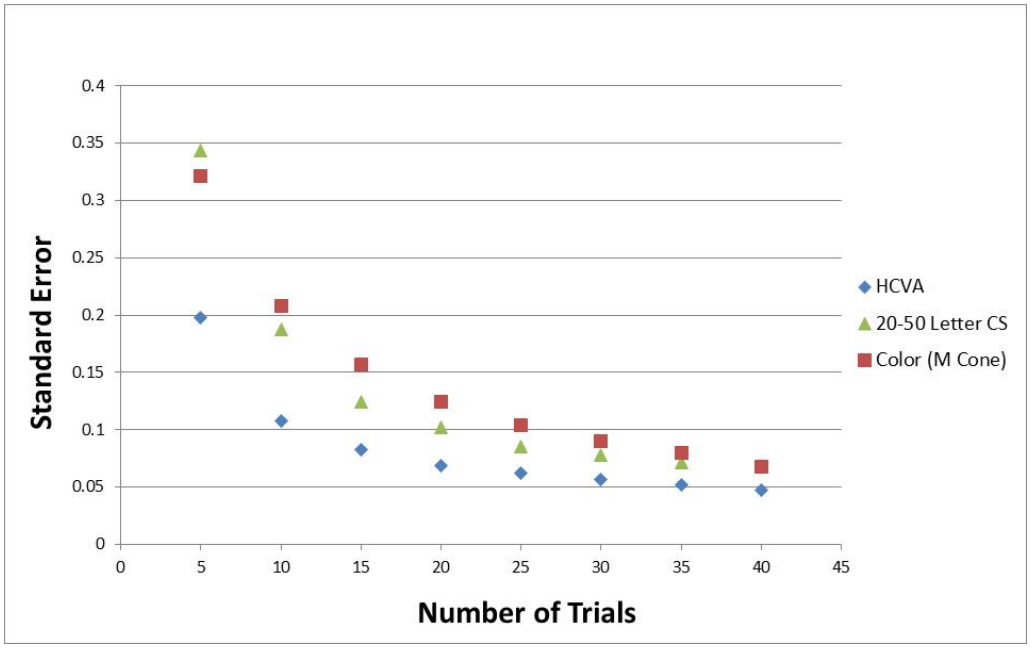

Figure 5. Estimated standard error on three visual tasks based on the number of trials completed.

Generally, the R^2 achieved after 40 trials on a given task was related to the level of homogeneity within the resultant data set. Tests with low correlations, e.g., high contrast VA, spanned a range of less than one log unit from the best to worst performers, and the majority of the results fell within a range of less than half of a log unit. Alternatively, tests with higher levels of correlation, e.g., 20/50 letter CS, spanned a range of up to two log units as shown in Figure 6.

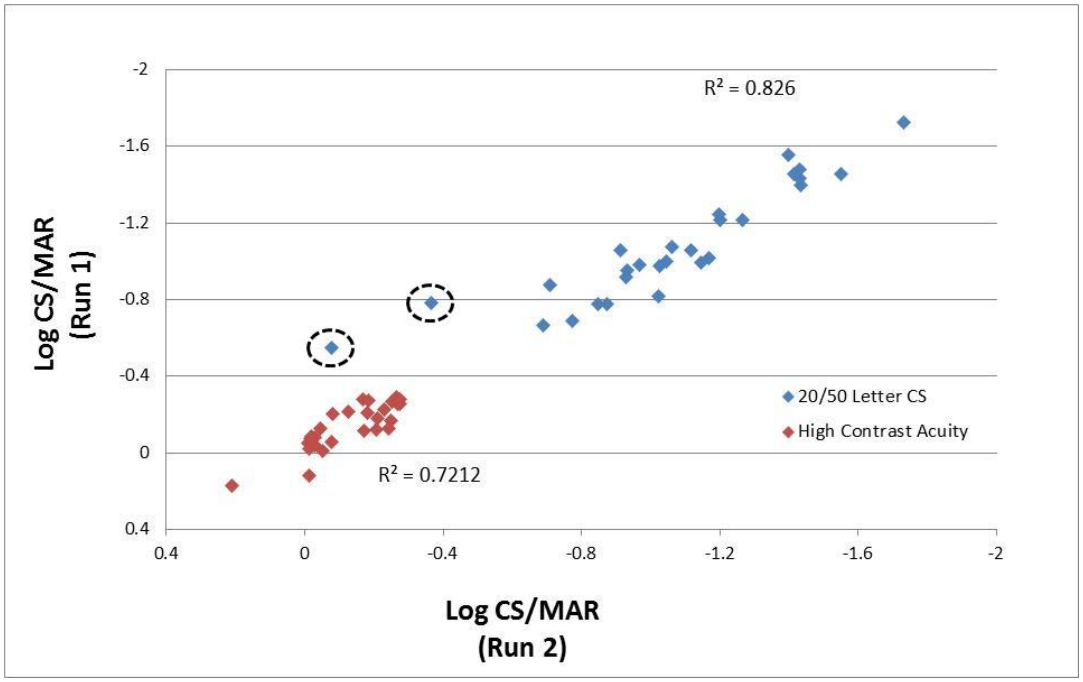

Figure 6. Comparison of test-retest repeatability between high VA and 20/50 letter CS.

A second method used to assess repeatability was to calculate mean differences between test one and test two as well as the standard deviation (SD) of these differences. This is reported in Table 2. Overall, mean differences between runs were low, although there was a small bias toward better performance on the second run. This implies a learning effect equating to an improvement of approximately 3.5% on the second test. Contrary to what was observed with correlation data, high contrast VA had relatively low variance on mean differences, while 20/50 letter CS had the highest variance. The latter finding was primarily due to two outliers circled in Figure 6.

Comparisons between the automated tasks and the analogous task using current methods (e.g., eye chart, RCCT, Titmus) are reported in Table 3. Results for the automated tasks were taken as the average threshold measurement on two runs after 40 trials, with the exception of stereoacuity, which was limited to 30 trials. Stereoacuity software was written prior to the other tasks, and this discrepancy was not noted until after data collection had been accomplished.

Table 2. Mean Threshold, in Log Units, for Test One and Test Two, Difference in Mean Threshold Between Test One and Test Two, and SD of the Difference for Each Automated Task

Task	Test One Mean	Test Two Mean	Difference	SD
High Contrast VA	-0.129	-0.137	-0.009	0.066
Low Contrast VA	0.358	0.332	-0.025	0.059
20/50 Letter CS	-1.062	-1.076	-0.015	0.152
20/25 Letter CS	-0.935	-0.950	-0.014	0.128
Color M Cone	-1.832	-1.839	-0.007	0.078
Color L Cone	-1.983	-1.983	0.001	0.090
Color S Cone	-0.980	-1.026	-0.046	0.094
Stereoacuity	1.292	1.282	-0.011	0.131

Table 3. Mean (SD) Log Thresholds for Automated and Manual Tasks

Task	Automated Task	Current "Manual" Task	R^2
High Contrast VA	-0.133 (0.119)	-0.111 (0.093)	0.627
Low Contrast VA	0.345 (0.169)	0.239 (0.140)	0.376
20/50 Letter CS	-1.069 (0.332)	-1.444 (0.189)	0.169
20/25 Letter CS	-0.931 (0.316)	-1.035 (0.291)	0.438
Color M Cone	-1.835 (0.308)	-1.763 (0.237)	0.855
Color L Cone	-1.983 (0.222)	-1.846 (0.141)	0.691
Color S Cone	-1.003 (0.173)	-0.781 (0.034)	0.050
Stereoacuity	1.542 (0.295)	1.437 (0.258)	0.714

Several findings are observed from this data set:

- On tasks that are currently administered using eyecharts (acuity, CS) or booklets (stereoacuity), subjects had a higher (poorer) threshold on the automated tasks.
- Subjects had a lower (better) threshold on color testing for all cone types using the OCCT due to a ceiling effect on the RCCT as shown in Figure 7.
- Variances were larger for every visual task when performed under automated conditions, particularly evident on the 20/50 letter CS task shown in Figure 8.

These findings will be discussed in greater detail later.

Until this point, all of the findings reported were based on threshold estimates established after a relatively large number of trials. From a clinical standpoint, due to time constraints involved with screening large numbers of subjects, it would not be practical to field a test that requires 30 or 40 trials on multiple visual tasks to determine if a subject met the established criteria. The ability to perform rapid screenings, in the absence of measuring threshold, is necessary. For this purpose, we evaluated two screening methods. The first screening strategy involved eight presentations of each visual stimulus at a level corresponding to the current pass/fail criteria used by the USAF as reported in Table 4. Eight was chosen as we felt it provided the minimum number of presentations necessary to afford an acceptably low risk of

passing the screening by sheer chance while allowing for one finger error. With eight presentations, the probability of offering at least seven correct responses by randomly guessing is 0.04%. In contrast, reducing the number of screening stimuli to five would increase the probability to 1.6%. The second screening strategy evaluated was to classify the subject as normal vs. abnormal based on the threshold achieved after eight trials using the same adaptive method described above.

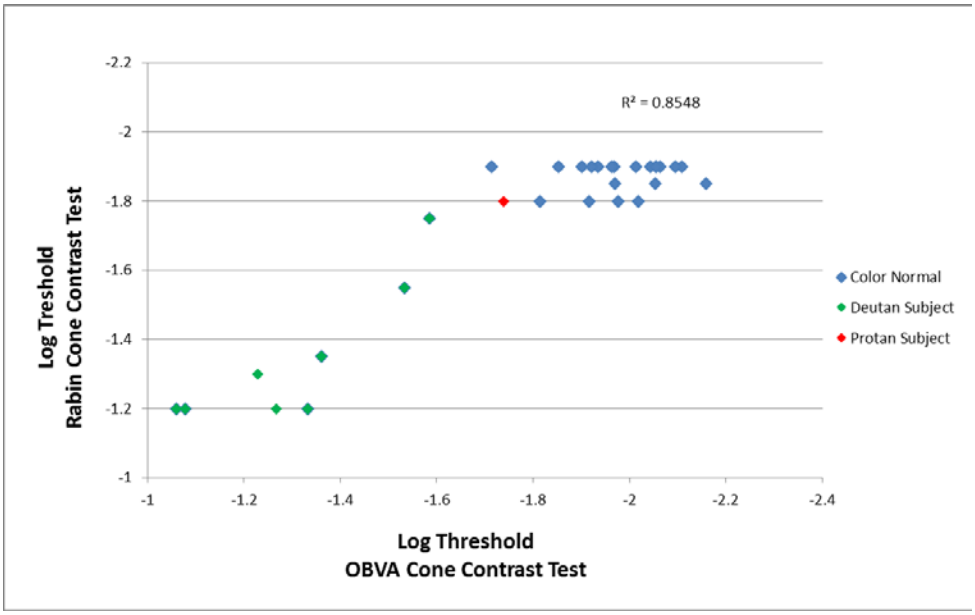

Figure 7. Comparison of M cone threshold for current RCCT and OCCT.

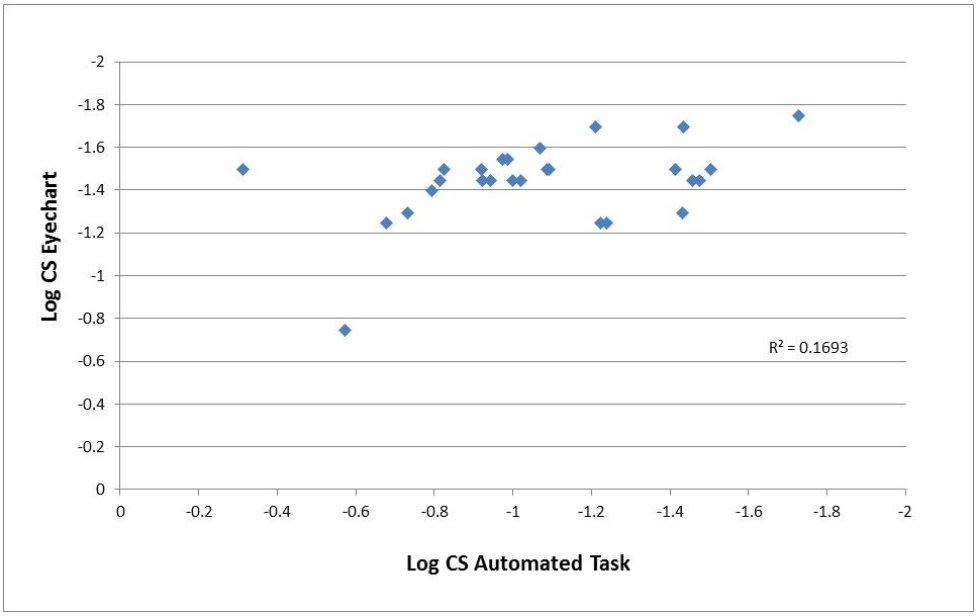

Figure 8. Comparison of threshold CS for automated task and eyechart for 20/50 letter CS.

Table 4. Pass/Fail Screening Characteristics for Each Automated Visual Task

Task	Pass/Fail Criteria	
	Log Units	Conventional Units
High Contrast VA	0.00 LogMAR	20/20 Snellen
5% Contrast VA	0.40 LogMAR	20/50 Snellen
20/50 Letter CS	-1.40 log CS	4.0% contrast
20/25 Letter CS	-0.80 log CS	15.8% contrast
Color M Cone	-1.66 log CS	2.2% contrast
Color L Cone	-1.66 log CS	2.2% contrast
Color S Cone	-0.55 log CS	28% contrast
Stereoacuity	1.40 log arcsec	25 arcsec

Results from the screening tasks were related to the final threshold obtained after 40 trials (30 for stereo). Specificity was defined as the number of subjects who passed the screening test (i.e., correctly identified at least seven of the eight screening stimuli and met or exceeded the passing score after eight trials) and met or exceeded the passing score for the given task after 40 trials. Similarly, sensitivity was defined as the number of subjects who failed the screening and whose threshold was below the pass/fail criteria after 40 trials. These results are reported in Tables 5 and 6.

Table 5. Screening Characteristics when Estimating Threshold from Eight Trials

Characteristic	VA		CS		Color Vision			Stereoacuity	All Tasks Combined
	High Contrast	5% Contrast	20/50 Letter	20/25 Letter	M Cone	L Cone	S Cone		
Specificity	96% (51)	94% (47)	68% (41)	89% (47)	100% (42)	98% (52)	94% (54)	98% (44)	93% (378)
Sensitivity	67% (3)	71% (7)	85% (13)	86% (7)	92% (12)	100% (2)	NA (0)	75% (10)	82% (54)

Note: Values in parentheses indicate number of subjects for each task.

Table 6. Screening Characteristics when Requiring at Least Seven of Eight Correct Responses on Screening Stimuli

Characteristic	VA		CS		Color Vision			Stereoacuity	All Tasks Combined
	High Contrast	5% Contrast	20/50 Letter	20/25 Letter	M Cone	L Cone	S Cone		
Specificity	75% (51)	81% (47)	71% (41)	77% (47)	90% (42)	98% (52)	87% (54)	89% (44)	84% (378)
Sensitivity	100% (3)	86% (7)	92% (13)	100% (7)	100% (12)	100% (2)	NA (0)	100% (10)	96% (54)

Note: Values in parentheses indicate number of subjects for each task.

Basing the screening result on the threshold achieved after eight trials yielded a specificity of 93% and a sensitivity of 82%. Presenting eight stimuli at the minimum passing criteria and requiring seven corrected responses yielded a specificity of 84% and a sensitivity of 96%. If a sensitivity of less than 100% was considered unacceptable, one could require subjects to correctly identify all eight of the screening stimuli. As shown in Table 7, this achieved the goal of properly identifying every subject performing below passing standards, however, at the cost of performing full threshold measurements on an additional 15% of normal subjects.

Table 7. Screening Characteristics when Requiring Eight Correct Responses on Screening Stimuli

Characteristic	VA		CS		Color Vision			Stereoacuity	All Tasks Combined
	High Contrast	5% Contrast	20/50 Letter	20/25 Letter	M Cone	L Cone	S Cone		
Specificity	55% (51)	68% (47)	49% (41)	55% (47)	79% (42)	88% (52)	70% (54)	82% (44)	69% (378)
Sensitivity	100% (3)	100% (7)	100% (13)	100% (7)	100% (12)	100% (2)	NA (0)	100% (10)	100% (54)

Note: Values in parentheses indicate number of subjects for each task.

5.0 DISCUSSION

The current effort demonstrates that automated vision tests produce reliable results, with coefficients of determination for repeated testing above 0.80 for many of the tasks. However, to achieve this high level of reproducibility and reduce the standard error of the threshold estimate to near asymptotic levels, 30 or more trials are often required.

Correlation of the automated tests to the current methods of administration produced more modest results. Only two tasks (M cone color testing and stereoacuity) produced R^2 values above 0.70, while S cone color resulted in an R^2 of 0.05. S cone correlation was particular low due a combination of the ceiling effect described with the RCCT and the fact that no subjects had an S cone (tritan) deficiency.

Comparison of automated tests developed for this study relative to counterpart tests using current methods yielded several findings of note:

- On tasks that are currently administered using eyecharts (acuity, CS) or booklets (stereoacuity), subjects had higher (poorer) thresholds on the automated tasks. This may be related to the fact that the current methods do not restrict viewing time, which potentially allows subjects to scan the visual stimulus and gain information that may not be available when the viewing time is restricted. A second explanation is that the score for the eyecharts is derived from the number of letters successfully identified without penalty for error, whereas errors on automated tasks drive the estimate of the threshold higher.
- Subjects had lower (better) thresholds on color testing for all cone types using the RCCT. This is almost certainly due to the fact that the RCCT has a ceiling effect that is not observed with the OCCT, as demonstrated in Figure 7.
- Variances were larger for every visual task when performed under automated conditions, particularly evident on the 20/50 letter CS task shown in Figure 8. Review of the data

collected from the chart shows that 17 of the 27 subjects had a threshold measured between -1.35 and -1.55 log CS, which represents a single line on the chart. In contrast, when evaluated with the automated task, these same 17 subjects had thresholds ranging from -0.80 to -1.50 log CS. Given that this automated task was proven to be highly reliable (Table 1), this suggests that the eyechart is not sensitive to small differences in performance between subjects.

We evaluated several screening techniques as an effort to maximize testing efficiency. When the status of a subject was based on the threshold estimate after eight trials, the specificity was high (93%), but the sensitivity was reduced (82%). We also implemented a set of eight screening stimuli set at the pass/fail criteria for each visual task and required at least seven correct responses for a passing score. This yielded a specificity of 84% and a sensitivity of 96%. Thus, 84% of normals could be confirmed with eight presentations, while the remaining 16% of normals required full threshold testing to properly categorize their visual status. Higher sensitivity could be achieved by requiring subjects to properly identify all eight of the screening stimuli. This achieved 100% sensitivity at the cost of performing full threshold testing on an additional 15% of normal subjects.

6.0 CONCLUSIONS

Our results suggest that automated vision testing can be successfully implemented. Although this effort was accomplished using desktop monitors, future efforts will pursue transitioning these tests to a system designed to standardize test distance and illumination conditions and eliminate the potential for head movements, and with a form factor suitable for more routine clinical use.

Successful implementation of automated (or any) vision testing in an aerospace medicine clinic requires methods of determining visual status quickly, but with high accuracy. We found that 100% sensitivity could be achieved, however, at the cost of performing full threshold testing on over 30% of normal subjects, which is quite time consuming.

We offer two proposed explanations for this relative high rate of false positive findings. First, the pass/fail criteria applied to the computerized tests were established based on data collected from prior studies using the manual techniques (e.g., eyecharts, stereo book) and were defined as two SDs below mean levels for a normal population. However, in all cases (except high contrast acuity), mean automated results were poorer than those obtained with manual techniques, and the distribution (SD) was greater for all tasks with automated testing. Therefore, the pass/fail criteria used for automated testing were, in effect, more challenging than when applied to manual conditions. It likely represented near threshold limits for a number of subjects, and screening at threshold limits is problematic and inconsistent. A second possible explanation is finger errors during both the screening and threshold phase of testing. We used a keypad to capture responses, which is not a common interface used by young adults. It was proposed that a joystick or game controller may be more appropriate. These will be assessed in future studies.

7.0 REFERENCES

1. Moffitt K, Genco LV. Criteria for a state-of-the-art vision test system. Wright-Patterson AFB (OH): Air Force Aerospace Medical Research Laboratory; 1985. Technical Report No. AFAMRL-TR-85-004.
2. Armed Forces-NRC Vision Committee. The Armed Forces-National Research Council Vision Committee: history 1944-1949 and description of organizational structure. Ann Arbor (MI): Armed Forces-NRC Vision Committee Secretariat; 1951.
3. Watson AB, Pelli DG. QUEST: a Bayesian adaptive psychometric method. Percept Psychophys. 1983; 33(2):113-120.
4. Kontsevich LL, Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. Vision Res. 1999; 39(16):2729–2737.
5. Treutwein B. Adaptive psychophysical procedures. Vision Res. 1995; 35(17):2503-2522.
6. Gaska J, Winterbottom M, van Atta A. Operational Based Vision Assessment cone contrast test: description and operation. Wright-Patterson AFB (OH): U.S. Air Force School of Aerospace Medicine; 2016. Special Report No. AFRL-SA-WP-SR-2016-0007.
7. Rabin, J, Gooch J, Ivan D. Rapid quantification of color vision: the cone contrast test. Invest Ophthalmol Vis Sci. 2011; 52(2):816-820.
8. Rabin J. Quantification of color vision with cone contrast sensitivity. Vis Neurosci. 2004; 21(3):483-485.
9. Georgeson MA, Freeman TC, Scott-Samuel NE. Sub-pixel accuracy: psychophysical validation of an algorithm for fine positioning and movement of dots on visual displays. Vision Res. 1996; 36(4):605-612.
10. Ferwerda JA, Greenberg DP. A psychophysical approach to assessing the quality of antialiased images. IEEE Comput Graph Appl. 1988; 8(5):85-95.

LIST OF ABBREVIATIONS AND ACRONYMS

CS	contrast sensitivity
OCCT	Operational Based Vision Assessment cone contrast test
OVT	Optec Vision Test
R^2	coefficient of determination
RCCT	Rabin cone contrast test
SD	standard deviation
USAF	U.S. Air Force
VA	visual acuity